Augmented Reality Interfaces

Mona Singh • *Independent Consultant*

Munindar P. Singh • *North Carolina State University*

Technological advances, exploding amounts of information, and user receptiveness are fueling augmented reality's (AR) rapid expansion from a novelty concept to potentially the default interface paradigm in coming years. This article briefly describes AR in terms of its application in natural Web interfaces. The authors discuss key concepts involved in AR, and the technical and social challenges that remain.

Augmented reality (AR) user interfaces have improved tremendously in the past few years. AR is drawing considerable interest not only because it involves novel or "cool" technologies, but also because it promises to help users manage today's information overload. AR helps present information succinctly, in its "natural" home, where users can easily benefit from and act on it.

We define AR as follows: AR presents a view of the real, physical world that incorporates additional information to augment this view. Of course, all the ways we view the world are just that — views. An implicit intuition is that the first view is somehow direct or canonical in that we can treat it as reality and further augment it with additional information. Such augmentations are "information" in the broadest sense, might include nonsense or false information, and can express any data type (text, image, video, and so on). A baseline example according to our definition would be a bird's eye view or satellite picture of a city (the "reality") overlaid with street and building names (the "augmentation").

AR is most naturally associated with settings in which the aspect of reality considered is current and proximal to the user; the augmenting information can likewise be current and proximal, or not, depending on specific settings. Moreover, the most common settings involve visual representation (whether still images or videos), although, in principle, we can augment any interface modality. For example, an app might play audio signals from the environment along with commentary on the relevant sounds (such as bird calls for ornithologists or various safety warning chimes for building occupants in training).

In this short article, we focus on how AR — especially phone-based AR, which is becoming widely available — can help provide natural Web interfaces.

Augmented Reality Examples and Nonexamples

Several AR apps are currently available:

- *Navigation*. Such an app would highlight the directions a user is taking — for instance, stating whether a turn is coming up. Vehicular displays might identify the appropriate highway lane or next turn. Figure 1 shows a screenshot of an Android AR navigation app (https://play.google.com/store/apps/details?id=com.w.argps).
- *Commerce*. A common theme is presenting advertisements according to the user's location or, more specifically, for any object recognized in a camera view. Figures 2a and 2b show how the Blippar app (http://blippar.com) progresses, beginning with the user pointing a phone camera at a grocery item. The app first recognizes the real-world object

Published by the IEEE Computer Society

1089-7801/13/$31.00 © 2013 IEEE

(bottle), and then augments it with an interactive object (recipe book).

- *Captioning*. Generalizing from Blippar, a user would point a phone camera at a scene. The phone would display a real-time image augmented with metadata associated with the scene or its salient parts. For example, a user could point a camera at a remote mountain peak and see its name, height, and current weather. Alternatively, the app might identify landmarks in a city or provide category descriptions (such as "restaurant" or "museum") for various buildings. Nokia City Lens is one such app (www.1800pocketpc.com/nokia-city-lens-augmented-reality-location-app-for-lumia-devices/).

Additional examples involve augmenting art, educational and gaming content, and fashion. An example in fashion is showing how a user would appear when wearing specified apparel.

Although our definition of AR is broad, it excludes certain applications, even though some might describe them as AR:

- *Immersive virtual reality (IVR)*. AR exposes the real world to a user with virtual information embedded in it, whereas IVR places a user in a virtual world (see, for example, www.kinecthacks.com/augmented-reality-telepresence-via-kinect/).
- *Photo editing*. One example is Mattel's "digital mirror," wherein users can edit pictures of themselves with cosmetics (http://mashable.com/2013/02/11/barbie-makeup-mirror/). Another is the Snaps iPhone app, which enables users to add images of celebrities to a photo (https://itunes.apple.com/us/app/snaps!/id600868427?mt=8). These apps don't augment reality, although using the edited pictures in place of users' original faces in a real scene might be considered a form of authoring content for an AR app that augments reality with the edited pictures.
- *Augmented media*. An example is the *Guinness Book of World Records*, which recently provided 3D animations of some world records (www.appsplayground.com/apps/2012/09/03/augmented-reality-sharks-star-in-guinness-world-records-2013-app/).

Editor's Introduction

Advances in consumer devices and back-end infrastructure are changing how we interact with information on the Web. This department, Natural Web Interfaces, will present a series of articles on recent or emerging technologies, methods, and standards for how users engage with the Web, emphasizing approaches that bring out the "naturalness" of the user experience. Relevant themes include diverse modalities such as speech and touch, interaction with the social and physical worlds, and a treatment of challenges such as architecture and engineering. —*Munindar Singh*

Figure 1. Augmented reality GPS Drive/Walk Navigation app. The app augments the user's view of reality with the suggested path, indicating the distance to the next turn.

The distinction between AR and what we term *augmented media* falls along a continuum. We can imagine pure AR as the augmentation of "natural" reality. However, all too often, AR works only when reality has been suitably prepped. An example is the Amazon app (www.amazon.com/gp/anywhere/sms/android). Here, the user takes a picture of a product's barcode, and the app finds that product on Amazon and presents a user interface to enable immediate purchase. The app relies on a media object — the barcode — that would be embedded in the product without regard to AR. Going further, we could affix Quick Response (QR) codes on physical artifacts specifically for AR (see, for example, www.npr.org/2013/07/29/206728515/activists-artists-fight-back-against-baltimores-slumlords), in effect treating the reality as less natural and more symbolic. As with the Guinness Book example, the extreme form occurs when the user interaction occurs purely with the media object and has no direct bearing on the reality.

Figure 2. Blippar. The app (a) recognizes the real-world object (bottle), and then (b) augments it with an interactive object (recipe book) identifying potential uses of the product.

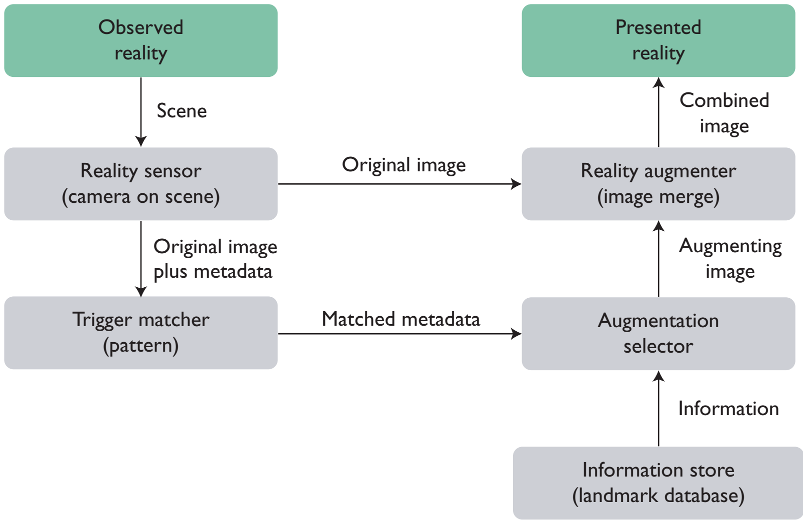

Figure 3. Conceptual architecture of an augmented reality app. An AR app resides on the path from reality as observed to reality as presented (with augmentations) to a user. It involves determining the relevant part of the reality, finding whether and how to augment it, and then inserting the augmentation and presenting the augmented view to a user.

AR Architecture

Figure 3 shows a conceptual reference architecture for an AR app, including its essential components and some image-related annotations as examples. (AR could potentially apply to any sense, including hearing.) A *reality sensor* (camera) observes a part of the reality. It then passes the image it obtains along with metadata such as geolocation tags to the *trigger matcher*, which checks if its input matches the relevant app-specific trigger. Such triggers could include geolocation that's near a specific landmark or the image showing the landmark. The trigger matcher then produces matched metadata, such as the image's semantic category and outline. The *augmentation selector* takes the matched metadata from the trigger matcher and retrieves relevant information, such as the year the landmark was built. It constructs an augmenting image, such as a text bubble or a map pin that can be placed relative to the original image, and passes it to the *reality augmenter*, which combines the images and renders them for the user. The same structure would apply if we think not of images but of video streams. And the architecture would often be enhanced with other modules to more naturally determine what element of the scene is most relevant to a user and how a user could interact with the augmented view — for example, by tracking the user's gaze.

Realizing Augmented Reality

Realizing AR requires high-quality sensing, computing, and communications platforms, but not more so than are becoming common today.

Enabling Technologies

The architecture we've described highlights the necessary enabling technologies for realizing AR.

To obtain a sufficiently accurate representation of reality, AR first needs suitable sensors in the environment and possibly on a user's person, including fine-grained geolocation and image recognition. Second, trigger matching and image augmentation require means to understand the scene before they can determine the relevant components and display augmentations — for instance, through techniques such as image processing (with face recognition being an important subcategory).

Third, trigger matching and subsequent user interaction presume ways to determine the user's attention and immediate context — for instance, via technologies for input modalities that include gaze tracking, touch, and gesture and speech recognition. Fourth, AR presupposes a substantial information infrastructure — for instance, accessible via cloud services — for obtaining pertinent components of the user's longer-term context. Such components include intent and activities. This infrastructure must also determine what real-world components to augment, with what, and when. Finally, AR requires significant computing and communications infrastructure undergirding all these technologies.

User Platforms

The aforementioned technologies are realized on three main types of end-user platforms, each against a cloud services backdrop. Mobile phones are the most prevalent of these platforms, with vehicles and wearable computers to soon follow. Modern phones include high-quality cameras, geolocation capabilities, numerous other sensors, and sufficient computing and communications capabilities.

A driver in a vehicle needs to access information about nearby and upcoming locations. A vehicle's windshield provides an intuitive venue for rendering augmented information. Vehicles have practically unlimited (electric) power and can support powerful computing and communications.

Wearable computers, such as the well-known Google Glass, are becoming viable. Like smartphones and vehicles, wearable computers provide numerous sensors and close access to users' current environment and their immediate context and attention. Wearable sensors, including on users' skin, clothing, or shoes, offer access to biometric and environmental data and can thus enable smart apps. Today's wearable computers are, however, limited in power, computing, and communications.

Toward an AR App Taxonomy

An AR app has several essential but wide-ranging ingredients, suggesting a classification of AR apps along the following dimensions.

The *trigger* is the event or observation based on which the augmentation occurs. Typical instances are location or object recognition (which could occur at multiple levels of granularity, ranging from object types to faces of specific people). One type of location trigger is matching on GPS coordinates. For example, Nokia City Lens provides information about nearby places of interest, letting users search for restaurants, hotels, and shops, and obtain more information about them.

Blippar (Figures 2a and 2b) exemplifies object recognition. Having a phone provide relevant information from a barcode is quite common. As mentioned, the Amazon Mobile app helps users obtain the product description from Amazon for any UPC symbol their camera captures. Similar apps are available from Google Shopper and eBay's RedLaser (https://play.google.com/store/apps/details?id=com.ebay.redlaser).

An example of face recognition is Recognizr, a now-defunct augmented ID app (www.tat.se/blog/tat-augmented-id/) that identifies a person and displays his or her online profile and contact details.

Interactivity is the extent to which users can interact with the augmented information through the app. In general, with apps that show reality in a direct view, the user might have occasion to interact only with the augmented information, not the reality. An example of no interactivity is road names augmented on a satellite image; an example of low interactivity is the GPS Drive/Walk app Figure 1 shows; Blippar is an example of medium interactivity, because users can request a recipe or video by selecting the appropriate marker. The BMW Service app (www.bmw.com/com/en/owners/service/augmented_reality_introduction_1.html) also exhibits medium interactivity: it displays servicing instructions and advances them whenever a user asks for the next step. An example of high interactivity is advertisement icons that open up automatically to reveal discounts when approached.

User interface modalities constrain how a user interacts with the augmented information through gesture, gaze, speech, and touch in addition to traditional modalities such as joysticks. Touch and speech are common these days. Google Glass provides a speech interface.

Finally, *naturalness of view* can trigger the AR app based on natural reality (Recognizr) or require specific features embedded in the environment or physical objects (Amazon).

Opportunities and Prospects

Modeling and applying user context remains the key challenge to realizing high-quality user experience. AR promises to present information and support user actions in ways that are sensitive to the user's current context.

Usability Challenges

AR faces the same core usability challenges as traditional interfaces, such as the potential for overloading users with too much information and making it difficult for them to determine a relevant action. However, AR exacerbates some of these problems because multiple types of augmentation are possible at once, and proactive apps run the risk of overwhelming users. Designers should focus on several key questions to address usability concerns.

Can the user tell the difference between reality and the augmentation? Confusion could lead to user errors if an app conveys an erroneous impression of the world.

Is the augmentation aligned with reality? Maintaining alignment is nontrivial because reality can change fast, especially in unanticipated ways. For example, in an AR navigation app, the traffic signal might change state, or an accident could occur well before the augmented information is updated.

How can a user transition between AR and traditional apps? For example, a user searching for a product might need to move between an AR-enabled app (to identify relevant products) and a traditional one (to search and purchase). However, transitions across apps could be confusing if their underlying metaphors are incompatible.

How should the augmenting information be organized? For example, if a relevant product comes in different varieties, colors, or prices, it would help to group related products in a way that's aligned with the user's intent. An AR app that presents all the information at once might serve only to confuse and mislead the user.

Social Challenges

AR is strikingly different from previous computing technologies in terms of both what it accomplishes and its physical trappings. As with other new technologies, however, it might take years before people begin to widely adopt it except in settings where there's a pressing need or a significant immediate benefit. We speculate that navigation and tourism in consumer usage and field maintenance in enterprises would be popular settings for AR.

Because AR is most useful when the augmentations are salient given the user's context — including attributes and prior experiences — privacy violations of users or those nearby is a potential risk. For example, an advertisement would be most useful if it were for something the user wanted. However, on receiving such an effective advertisement, users might wonder about how their personal information has propagated across the value chain.

Business Models

From a business model standpoint, we anticipate that AR apps would function like traditional ones in many respects. A key difference would regard who owns — that is, controls — the AR space. Presumably, the current app (or the entity that controls it) would control the display. For instance, instead of advertisements being displayed for keywords, as with today's Web, advertisements in AR might be displayed for appropriate triggers, such as particular locations or patterns. However, just this seemingly technical change from keywords to locations or patterns could lead new entities to emerge in the business ecosystem, such as those that would tackle maintaining the augmented information.

The AR apps we see today are little more than proofs of concept, but they do succeed in showing the potential of making augmented reality a reality. As technology improves over time, we expect AR to become a routine form of user experience.

Acknowledgments

Thanks to Pankaj Mehra for comments and to the US National Science Foundation for partial support for Munindar Singh under grant 0910868.

Mona Singh is an independent consultant specializing in user experience and innovative technologies for user interaction and has worked extensively on these topics. Singh has a PhD in linguistics from the University of Texas at Austin. Contact her at mona.singh.n@gmail.com.

Munindar P. Singh is a professor at North Carolina State University. His research interests include multiagent systems and context-aware computing. Singh has a PhD in computer sciences from the University of Texas at Austin. He's an IEEE Fellow. Contact him at m.singh@ieee.org.